

## METHOD AND DEVICE FOR GIVING TAG TO INFORMATION STRING AND RECORDING MEDIUM RECORDED WITH THE METHOD

**Publication number:** JP2000268033 (A)

**Publication date:** 2000-09-29

**Inventor(s):** SASAKI YUTAKA; HIROTA KEIICHI +

**Applicant(s):** NIPPON TELEGRAPH & TELEPHONE +

**Classification:**

- international: G06F17/21; G06F17/24; G06F17/27; G06F17/30; G06F17/21; G06F17/24;  
G06F17/27; G06F17/30; (IPC1-7): G06F17/27; G06F17/24; G06F17/30

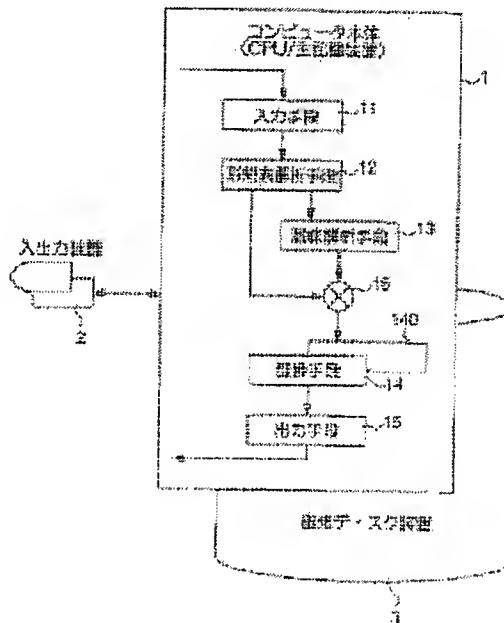
- European:

**Application number:** JP19990067562 19990312

**Priority number(s):** JP19990067562 19990312

### Abstract of JP 2000268033 (A)

**PROBLEM TO BE SOLVED:** To easily deal with the specification change or addition of a tag by automatically giving the tag to an information string and easily identifying the replacement of giving a certain specified tag. **SOLUTION:** An input means 11 fetches the information string through an input/output device 2. A morpheme analytic means 12 adds morpheme information to each word included in the inputted information string. A semantic analytic means 13 gives a semantic category to the inputted word and morpheme information string. A replacing means 14 gives a tag by performing replacement to the information string consisting of words and morphemes or information string consisting of words, morphemes and semantic category. On the basis of a series of simple replacement rules for giving one tag to the inputted information string, the information string can be tagged and the respective simple and independent replacement rules are added, deleted and changed to deal with the specification change or addition of the tag.



Data supplied from the **espacenet** database — Worldwide

(19)日本国特許庁 (JP)

## (12) 公開特許公報 (A)

(11)特許出願公開番号

特開2000-268033

(P2000-268033A)

(43)公開日 平成12年9月29日(2000.9.29)

(51)Int.Cl. <sup>7</sup>	識別記号	F I	テマコード*(参考)
G 0 6 F 17/27		G 0 6 F 15/38	E 5 B 0 0 9
17/24		15/20	5 5 0 E 5 B 0 7 5
17/30			5 5 4 H 5 B 0 9 1
		15/40	3 7 0 A
		15/401	3 1 0 A

審査請求 未請求 請求項の数9 O.L (全 6 頁)

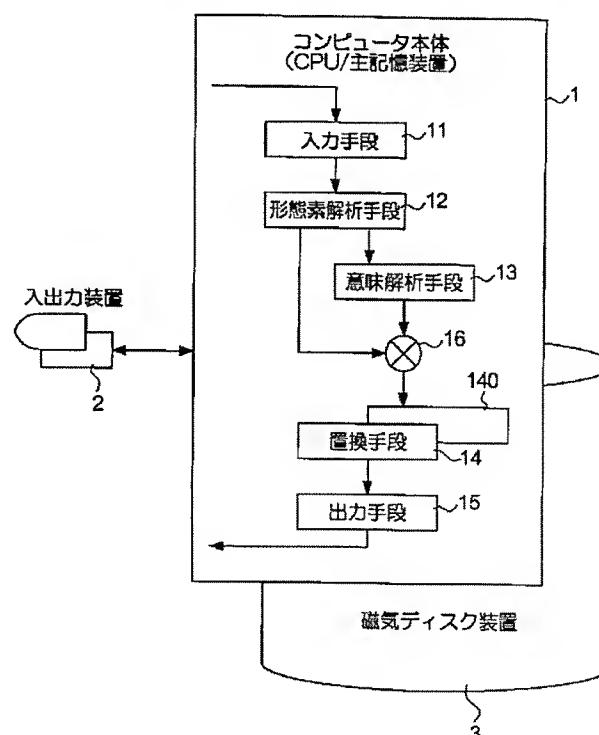
(21)出願番号	特願平11-67562	(71)出願人	000004226 日本電信電話株式会社 東京都千代田区大手町二丁目3番1号
(22)出願日	平成11年3月12日(1999.3.12)	(72)発明者	佐々木 裕 東京都新宿区西新宿三丁目19番2号 日本 電信電話株式会社内
		(72)発明者	廣田 啓一 東京都新宿区西新宿三丁目19番2号 日本 電信電話株式会社内
		(74)代理人	100064908 弁理士 志賀 正武 Fターム(参考) 5B009 QA09 QB14 QB18 5B075 ND03 NK32 5B091 AA15 CA02

(54)【発明の名称】 情報列に対してタグ情報を付与するための方法および装置ならびに同方法が記録される記録媒体

## (57)【要約】

【課題】 情報列に対してタグを自動付与し、ある特定のタグを付与する置換を容易に識別でき、タグの仕様変更や追加に容易に対応する。

【解決手段】 入力手段11は、入出力装置2を介して情報列を取り込む。形態素解析手段12は入力された情報列に含まれる各単語に形態素情報を付加する。意味解析手段13は、入力された単語、形態素情報列に対して意味カテゴリを付与する。置換手段14は単語、形態素から成る情報列、あるいは、単語、形態素、意味カテゴリから成る情報列に対し、置換を施してタグを付与する。入力情報列に対して1つのタグを付与する一連の単純な置換ルールによって情報列に対するタグ付けを実現し、個々の単純で独立した置換ルールを追加、削除、変更することによってタグの仕様変更や追加に追従する。



**【特許請求の範囲】**

**【請求項1】** 単語を含む情報列を取り込み、あらかじめ用意された単語辞書を索引することにより該情報列に含まれる各単語の形態素情報を得て、少なくとも前記単語、前記形態素情報の組から成る情報列に対し、あらかじめ定義された規則に従って置換を施すことにより前記単語を特定するタグを付与し、該タグを含む情報列を出力することを特徴とする情報列に対してタグ情報を付与するための方法。

**【請求項2】** あらかじめ用意された意味辞書を索引して前記単語について最適な意味カテゴリ情報を得て、前記単語、前記形態素情報、前記意味カテゴリ情報の組から成る情報列に対し、あらかじめ定義された規則に従って置換を施すことにより前記単語を特定するタグを付与することを特徴とする請求項1記載の情報列に対してタグ情報を付与するための方法。

**【請求項3】** 前記置換は、入力される情報列に対して1個のタグを付与する一連の置換ルールを用い、単語、形態素情報、あるいは、単語、形態素情報、意味カテゴリ情報から成る特定のパターンを、逐一、パターンマッチングによりタグが付与された情報列のパターンに変換することを特徴とする請求項1または2記載の情報列に対してタグ情報を付与するための方法。

**【請求項4】** 前記各置換ルールに名称を付与し、各置換ルールに従って付与されるタグに対して、該当する前記名称を付して履歴として残すことを特徴とする請求項3記載の情報列に対してタグ情報を付与するための方法。

**【請求項5】** 入力された前記情報列に対し、1個のタグを付与する一連の置換ルールを更新することにより、タグの仕様変更に追従することを特徴とする請求項1または2記載の情報列に対してタグ情報を付与するための方法。

**【請求項6】** 請求項1～5のいずれかの項に記載された方法の各手順をコンピュータに実行させるためのプログラムを記録したコンピュータ読み取り可能な記録媒体。

**【請求項7】** 入力される情報列を取り込む入力手段と、取り込まれた情報列に含まれる単語を形態素解析して得られる形態素情報を前記各単語に付加して出力する形態素解析手段と、前記情報列に対して1個のタグを付与する一連の置換ルールを使用して、少なくとも前記単語、前記形態素情報の組から成る情報列に対し、前記置換ルールのそれぞれに従ってタグ情報を付与する置換手段と、前記タグが付与された情報列を出力する出力手段とを有することを特徴とする情報列に対してタグ情報を付与するための装置。

**【請求項8】** 入力される情報列の各単語の意味を判定

し、前記単語及び前記形態素情報から成る情報列に対して意味カテゴリ情報を付加し、前記置換手段に対して、前記単語、前記形態素情報、前記意味カテゴリ情報の組からなる情報列を供給する意味解析手段を更に有することを特徴とする請求項7記載の情報列に対してタグ情報を付与するための装置。

**【請求項9】** 前記置換手段は、少なくとも2つの置換ルールを記憶する記憶装置を備え、該記憶装置から得られるそれぞれの置換ルールに従って置換を行ってタグ情報を付与することを特徴とする請求項7または8記載の情報列に対してタグ情報を付与するための装置。

**【発明の詳細な説明】****【0001】**

**【発明の属する技術分野】** 本発明は、コンピュータによる自然言語処理システムに用いて好適であって、情報列に対してタグ情報を付与するための方法および装置ならびに同方法が記録される記録媒体に関する。

**【0002】**

**【従来の技術】** コンピュータによる自然言語処理システムにおいて検索処理や置き換え処理を行うのに、通常、パターンマッチングの手法が使用される。このときに情報列へ付与されるタグは、従来、C言語やPerl（スクリプト言語）などのプログラミング言語を用いて、情報列の変換を行う一般的な情報処理プログラムを記述することによって生成していた。

**【0003】**

**【発明が解決しようとする課題】** しかしながら、上述したタグを付与するための処理は一般的のプログラムとして記述されているため、実用レベルのタグ付けを実現するためには、繰り返し処理や判断処理が複雑に錯綜しあうことにある。そのため、タグの種類の増減やタグの定義の変更に伴う装置変更の際の対応が複雑で煩わしく、保守性に乏しいものであった。

**【0004】** 本発明は上記事情に鑑みてなされたものであり、入力情報列に対して1つのタグを付与する一連の単純な置換ルールを複数直列に接続することによってタグ付けを実現し、ある特定のタグを付与する置換えを容易に識別できると共に、個々の単純で独立した置換ルールを追加、削除、変更することによってタグの仕様変更や追加に容易に対応でき、保守性の向上を図ることができる情報列に対してタグ情報を付与するための方法および装置ならびに同方法が記録される記録媒体を提供することを目的とする。

**【0005】**

**【課題を解決するための手段】** 上述した課題を解決するために、請求項1記載の発明は、単語を含む情報列を取り込み、あらかじめ用意された単語辞書を索引することにより該情報列に含まれる各単語の形態素情報を得て、少なくとも前記単語、前記形態素情報の組から成る情報列に対し、あらかじめ定義された規則に従って置換を施す

すことにより前記単語を特定するタグを付与し、該タグを含む情報列を出力することを特徴とする。また、請求項2記載の発明は、請求項1記載の発明において、あらかじめ用意された意味辞書を索引して前記単語について最適な意味カテゴリ情報を得て、前記単語、前記形態素情報、前記意味カテゴリ情報の組から成る情報列に対し、あらかじめ定義された規則に従って置換を施すことにより前記単語を特定するタグを付与することを特徴とする。また、請求項3に記載の発明は、請求項1または2記載の発明において、前記置換は、入力される情報列に対して1個のタグを付与する一連の置換ルールを行い、単語、形態素情報、あるいは、単語、形態素情報、意味カテゴリ情報から成る特定のパターンを、逐一、パターンマッチングによりタグが付与された情報列のパターンに変換することを特徴とする。

【0006】また、請求項4記載の発明は、請求項3記載の発明において、前記各置換ルールに名称を付与し、各置換ルールに従って付与されるタグに対して、該当する前記名称を付して履歴として残すことを特徴とする。また、請求項5記載の発明は、請求項1または2記載の発明において、入力された前記情報列に対し、1個のタグを付与する一連の置換ルールを更新することにより、タグの仕様変更に追従することを特徴とする。また、請求項6記載の発明は、請求項1～5のいずれかの項に記載された方法の各手順をコンピュータに実行させることを特徴とする。

【0007】また、請求項7記載の発明は、入力される情報列を取り込む入力手段と、取り込まれた情報列に含まれる単語を形態素解析して得られる形態素情報を前記各単語に付加して出力する形態素解析手段と、前記情報列に対して1個のタグを付与する一連の置換ルールを使用して、少なくとも前記単語、前記形態素情報の組から成る情報列に対し、前記置換ルールのそれぞれに従ってタグ情報を付与する置換手段と、前記タグが付与された情報列を出力する出力手段とを有することを特徴とする。また、請求項8記載の発明は、請求項7記載の発明において、入力される情報列の各単語の意味を判定し、前記単語及び前記形態素情報から成る情報列に対して意味カテゴリ情報を付加し、前記置換手段に対して、前記単語、前記形態素情報、前記意味カテゴリ情報の組からなる情報列を供給する意味解析手段を更に有することを特徴とする。また、請求項9記載の発明は、請求項7又は8記載の発明において、前記置換手段は、少なくとも2つの置換ルールを記憶する記憶装置を備え、該記憶装置から得られるそれぞれの置換ルールに従って置換を行ってタグ情報を付与することを特徴とする。

【0008】上述した本発明の構成において、入力手段は入力された情報列を取り込み、形態素解析手段は当該情報列に含まれる各単語に対して形態素情報を付加し、意味解析手段は、単語、形態素情報列に対して意味カテ

ゴリ情報を付与する。置換手段は、単語、形態素から成る情報列、あるいは、単語、形態素、意味カテゴリから成る情報列に対して置換を施することでタグを付与し、出力手段がタグ付き情報列を出力する。こうして、入力された情報列にタグを付与することができる。本発明によれば、入力情報列に対して1つのタグを付与する一連の単純な置換ルールにより情報列に対するタグ付けを実現することができる。これにより、ある特定のタグを付与する置換を容易に識別できるとともに、個々の単純で独立した置換ルールを追加、削除、変更することにより、タグの仕様変更や追加に容易に追従できる。

#### 【0009】

【発明の実施の形態】以下、図面を参照して本発明の一実施形態について説明する。図1は本発明の実施形態を示すブロック図である。図において、符号1はコンピュータ本体であり、CPU(中央処理装置)と主記憶装置を含む。符号2はキーボード・ディスプレイ等から成る入出力装置、符号3は辞書情報の他、プログラム乃至各種データが格納される磁気ディスク装置であって、大容量記憶装置の一例である。

【0010】コンピュータ本体1は、入力手段11、形態素解析手段12、意味解析手段13、置換手段14、出力手段15から成る。また、コンピュータ本体1は必要に応じてゲート手段16を有している。このゲート手段16は、形態素解析手段12および意味解析手段13から情報を得て、置換手段14に対してこれら両方の手段から得た情報を供給するか、もしくは、形態素解析手段12から得た情報のみを供給する。なお、符号140は置換ルール記憶装置であるが、これについては後で説明することにする。

【0011】上述した構成において、入力手段11は、情報列を入出力装置2から取り込んで形態素解析手段12に渡す。形態素解析手段12は、取り込まれた情報列に含まれる各単語に対して形態素情報を付加し、意味解析手段13に渡すか、あるいはゲート手段16を介して置換手段14に渡す。意味解析手段13は、形態素解析手段12から出力された単語、形態素情報列に対して、ここで生成される意味カテゴリ情報を付与して置換手段14へ渡す。置換手段14は、単語、形態素情報、意味カテゴリ情報から成る情報列に対して置換を施すことによってタグを付与し、出力手段15を介して入出力装置2へ渡す。このことによって、入出力装置2はタグ付き情報列を出力する。

【0012】次に、上記構成を用いて情報列にタグを付与するための動作を説明する。図2～図10は、図1に示した本実施形態の動作を説明するために引用した図である。このうち、図2は図1に示す本実施形態の動作手順をフローチャートで示した図である。また、図3は図2に示すフローチャートにおけるステップS24の置換え処理の詳細動作を示すフローチャートである。さら

に、図4～図8は具体的な情報列に対するタグ付与の例、形態素解析手段12による情報列の出力例、置換ルール記述S1の例、置換ルール記述S2の例、置換手段14による具体的変換例をそれぞれ示した図である。このほか、図9、図10は、それぞれ、置換ルールに従って付与されるタグに履歴を残す拡張記述の例を示す図、本実施形態による置換ルールの拡張記述の例を示す図である。

**【0013】**以下、図1に示す本実施形態の動作について、図2～図10を参照しながら詳細に説明する。なお以下では、ゲート手段16の作用によって、形態素解析手段12の出力が意味解析手段13に渡され、その後に、意味解析手段13の出力がゲート手段16を介して置換手段14に渡されるものとする。このほか、入力情報列としては、具体的には、図4(a)に示す情報列(1. 千葉、2. 総裁)が与えられ、人名を表すタグ<PERSON>と役職名を表すタグ<PTITLE>をそれぞれ付与して、図4(b)に示す結果(1. 千葉<PERSON>、2. 総裁<PTITLE>)を得る例について詳述する。

**【0014】**図4(a)において、単語「千葉」は人名と地名の可能性があるが、「千葉」が人名であることは人名タグ<PERSON>を付与することによって明確になる。なお、単語の前に付与した行番号“1”，“2”は説明を容易にするために付したものであって、これらは無くとも構わない。また、本実施形態では、説明を容易するために1個の単語を一行毎に分けて記述しているが、各単語が識別できればどのようなデータ表現形式を用いても構わない。例えば、各単語が\$記号で区切られた形式でもよい。これは、後述する説明で、品詞や意味カテゴリ、タグを付与しているデータ形式についても同様のことが言える。

**【0015】**まず、入力手段11は、入出力装置2を介して入力される情報列から単語列である「千葉」、「総裁」を取り込み、これらを形態素解析手段12へ渡す(図2ステップS21)。形態素解析手段12は、単語から品詞などの形態素情報を判定して当該単語に付加する(ステップS22)。形態素解析の手法は、例えば、「長尾 真 編：自然言語処理、岩波書店、1996発行」の文献に詳細に述べられている。基本的には、磁気ディスク装置3に格納され、必要に応じてコンピュータ本体1の主記憶装置にローディングされ使用される単語辞書から単語を探し、前後の単語から最適な品詞を選択することによって実現される。

**【0016】**ここでは、各単語の品詞につき、「千葉」は固有名詞、「総裁」は普通名詞と判定されたとする。これらの品詞は、単語の後に半角スペースを挿入し、図5に示すように付与される。そして、この結果は意味解析手段13に渡される。意味解析手段13は各単語に対して意味カテゴリを付与する(ステップS23)。なお、意味カテゴリとはその単語が意味している内容を表

すカテゴリである。例えば、「千葉」という単語には「人間」と「場所」の2つの意味があり、「総裁」には「役職」という意味があると考えられる。意味解析の手法は上述した同様の文献に述べられている。基本的に

は、各単語について意味辞書を索引して最適な意味カテゴリを選択することによって実現される。そして、意味解析手段13は意味カテゴリをさらに形態素情報の後に付加して置換手段14に渡す。尚、意味カテゴリは複数の可能性がありうるため、図8(a)に示すように角括弧で括ることにする。但し、本実施形態ではこの意味表現形式の制限を受けるものではない。

**【0017】**置換手段14は、入力に対して1つタグを付与する一連の単純な置換ルールからなる。本実施形態では、図6および図7にそれぞれ示される2つの置換ルールS1およびS2を考える。まず、役職を表すタグ<PTITLE>を付与する置換ルールS1は、図6に示すように記述される。この置換ルールS1は、図3にフローチャートで示すパターンマッチ(S242)を使った置換を表す。図6に示す“X”は変数であってどのような単語ともマッチする。意味解析手段13により出力される情報列の中に、図6中の矢印(↓)の上の行とマッチする行があった場合、矢印の下の行に置換える。この置換の際に“X”はマッチした元の単語と置換えられる。

**【0018】**次に、人名を表すタグ<PERSON>を付与する置換ルールS2は、図7に示すように記述される。この置換ルールS2は図3にフローチャートで示すパターンマッチ(ステップS243)を使った置換えを表す。図7において、“X”および“Y”は変数であってどのような単語ともマッチする。意味解析手段13により出力される情報列の中に、図7中の矢印(↓)の上にある連続した2行とマッチする行があった場合、矢印の下の2行に置換える。この置換の際に“X”および“Y”はマッチした元の単語と置換えられる。

**【0019】**尚、置換手段14に含まれる単純な置換ルールの表現形式は、別の表現形式をとっても構わない。例えば、入力された情報列の単語が\$記号で区切られている形式であれば、\$記号で区切られた単語、形態素、意味カテゴリの列の特定のパターンをタグを付与した列にパターン変換する置換ルールを用意することができる。

**【0020】**図3に示すフローチャートにおいて、置換手段14は、あらかじめ用意された置換ルールS1およびS2(ステップS241)に従って、意味解析手段13から出力される単語、形態素情報、意味カテゴリ情報の列に対して置き換えを行う。まず、置換ルールS1に従う置換を行なう(図3のステップS242)ことによって、情報列の2行目の単語「総裁」に対してタグ<PTITLE>が付与されて(ステップS246)、図8(a)に示す様に変換される。さらに、置換ルールS2(ステップS243)に従って、情報列の1行目の「千葉」の

品詞、意味カテゴリがタグ<PERSON>に置換されて（ステップS247）、図8（b）に示す結果が得られる。そして、この結果は置換手段14から出力手段15へ渡される（ステップS248）。出力手段15は、受取った情報（図8（b））を入出力装置2へ供給する。このようにして、入力情報列に対してタグが付与されて出力される。

【0021】尚、本実施形態におけるタグの付与に関する簡単な拡張の一つとして、置換手段14に2つ以上のタグを一度に付与する置換ルールをいくつか含めることがあげられる。また、上述した実施形態では、置換手段14が置換ルールと一体になっている例のみを示したが、置換ルールを別途用意される置換ルール記憶装置140（図1を参照）に記憶しておき、置換手段14が置換ルール記憶装置140から置換ルールを取りだしながら置換を行うように拡張することもできる。

【0022】また、保守性向上の観点から、置換手段14が持つ各置換ルールの名前を履歴として、その置換ルールが付与したタグに付加的につけておくことも考えられる。例えば、図9に示すように、置換ルールS1に従ってタグ<PTITLE>を付与する際には、置換ルールS1によってタグが付与されたことを示す履歴をコメント「#S1」のように付与する。尚、この場合、置換を行う際に履歴の部分は無視するようとする。履歴は特にコメントの形式に制限されるわけでもないし、別の記憶装置に記憶してもよい。これにより、最終結果に付けられたタグがどの置換ルールにより付与されたかをさらに容易に判定できるようになる。

【0023】置換手段14に含まれる置換ルールの記述も容易に拡張でき、例えば、変数を単語以外の部分に使うことができる。これにより、図10に示すように、意味カテゴリに拘わらず、品詞が「人名名詞」の場合には置換によってタグ<PERSON>を付与する置換ルールを記述することができる。

【0024】また、意味解析手段13が各単語の意味カテゴリを常に空とするという特殊なケースであっても、単語と品詞情報のみを使って本実施形態に示したタグ付けが可能である。このときには、ゲート手段16の作用によって、形態素解析手段12の出力が置換手段14に直接渡されてタグ付けされることになる。

【0025】以上説明したように、本実施形態では、入力情報列に対して1つのタグを付与する一連の単純な置換ルールを複数直列に接続することによってタグ付けを実現している。このため、ある特定のタグを付与する置換を容易に識別できると共に、個々の単純で独立した置換ルールを追加、削除、変更することによってタグの仕様変更や追加に容易に対応することができ、保守性の向上を図ることが可能となる。

【0026】尚、図1に示すコンピュータ本体1に含まれている入力手段11、形態素解析手段12、意味解析

手段13、置換手段14、出力手段15、ゲート手段16、および、図2～図3に示すフローチャートの処理はいずれもソフトウェアによって実現可能なものである。すなわち、当該ソフトウェアは、半導体記憶装置、磁気記憶装置、CD（コンパクト・ディスク）-ROM（読み出し専用メモリ）等の記録媒体に記録されて提供されるものであって、必要に応じてコンピュータ本体1にインストールされ、CPUがこれを読み出して実行することによって、上述した機能及び動作が実現されるものである。また、上述した実施形態では、コンピュータを用いた自然言語処理システムに用いられるものであるとしたが、この他に、情報検索システム、情報抽出システム等に用いても得られる効果が大きい。

[0027]

【発明の効果】以上説明のように、本発明によれば、単語を含む情報列に対しタグを付与することが可能となる。本発明では、1つのタグを付与する一連の単純な置換ルールによりタグが付与されるため、どのタグがどのような条件下で付与されるかが簡単に識別できる。このことにより、タグの付与の定義やタグの種類に変更があった場合に、容易にタグの付け方を変更でき、保守性が大幅に向向上する。

### 【図面の簡単な説明】

【図1】 本発明の一実施形態を示すブロック図である。

【図2】 図1に示す本実施形態の動作を説明するためのフローチャートである。

【図3】 図2に示すフローチャートにおけるステップS24の置換え処理の具体的動作手順を示したフローチャートである。

【図4】 同実施形態において、具体的な情報列に対するタグ付与の例を示す説明図である。

【図5】 同実施形態において、形態素解析手段12により出力される情報列を例示した説明図である。

【図6】 同実施形態において、置換ルール記述S1の例を示した説明図である。

【図7】 同実施形態において、置換ルール記述S2の例を示した説明図である

【図8】 同実施形態において、置換手段14による具体的な変換を例示する説明図である。

【図9】 同実施形態の置換ルールに従って付与されるタグに対して履歴を残す拡張記述の例を示した説明図である。

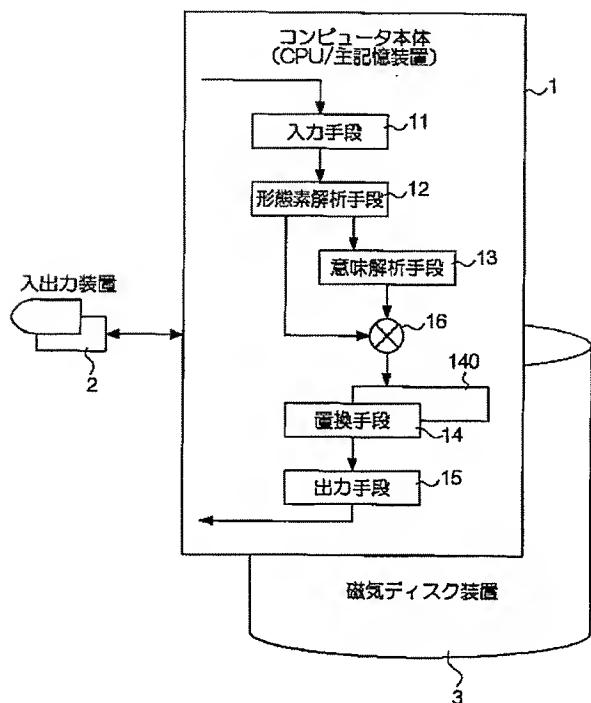
【図10】 同実施形態による置換ルールの拡張記述の例を示した説明図である

### 【竹見の説明】

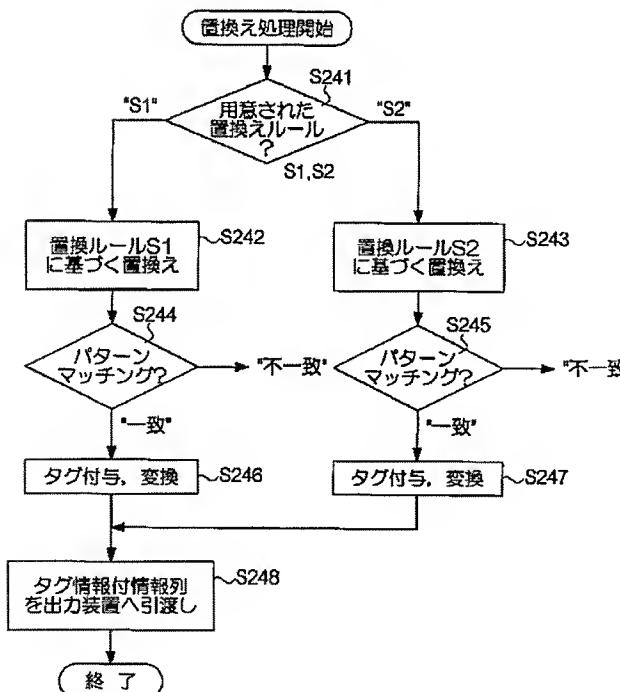
1…コンピュータ本体（CPU／主記憶装置）、2…入出力装置、3…磁気ディスク装置、11…入力手段、12…形態素解析手段、13…意味解析手段、14…置換手段、15…出力手段、16…ゲート手段、140…置換手段、141…意味表示手段、142…音声合成手段

## 換ルール記憶装置

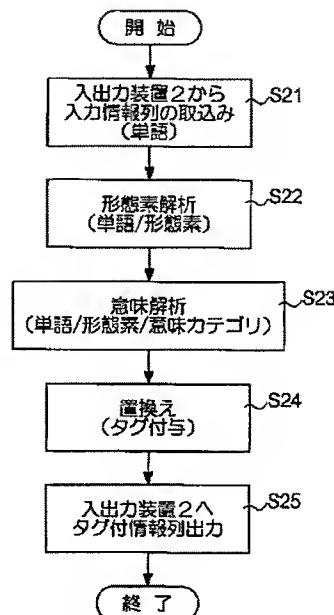
【図1】



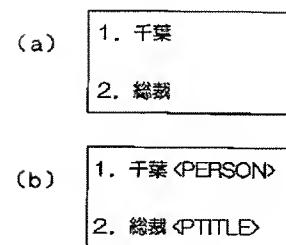
【図3】



【図2】



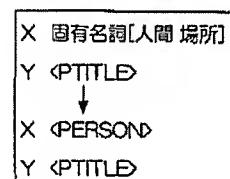
【図4】



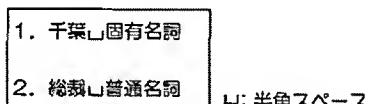
【図6】



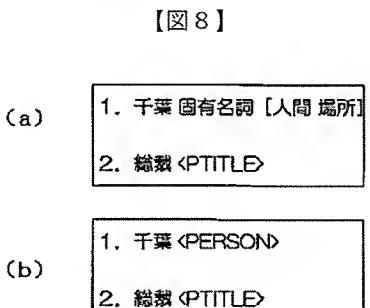
【図7】



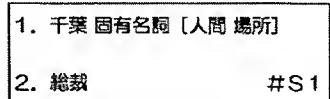
【図5】



【図10】



【図9】



【図8】